working paper
department
of economics

THE RELATIVE IMPORTANCE OF PERMANENT AND
TRANSITORY COMPONENTS;
IDENTIFICATION AND SOME THEORETICAL BOUNDS

Danny Quah

No. 498 June 1988

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

THE RELATIVE IMPORTANCE OF PERMANENT AND TRANSITORY COMPONENTS; IDENTIFICATION AND SOME THEORETICAL BOUNDS

Danny Quah

No. 498 June 1988

The Relative Importance of Permanent and Transitory Components:
Identification and Some Theoretical Bounds.

by

Danny Quah *
June 1988.

* Department of Economics, MIT and NBER. I am grateful to Olivier Blanchard and Jeffrey Wooldridge for ongoing discussions that have helped sharpen my understanding of the issues here. Conversations with Robert Engle and Mark Watson have also been useful. I thank the MIT Statistics Center for its hospitality. All errors and misinterpretations are mine.

The Relative Importance of Permanent and Transitory Components:
Identification and Some Theoretical Bounds.

by
Danny Quah
Economics Department, MIT.
June 1988.

Abstract

The relative contribution of permanent and transitory disturbances is a question of considerable importance in the study of economic fluctuations. A number of alternative empirical models have been proposed and used to estimate the relative sizes of these different components. These empirical models are typically just-identified or over-identified by assumptions that are open to dispute. This paper develops exact theoretical bounds on the relative importance of the permanent and transitory components focusing only on assumptions of orthogonality and lag lengths. The paper shows that the orthogonality restriction is inessential and that there is a direct relation between the theoretical minimum importance of the permanent component and assumptions on its lag length. Thus the importance of the permanent component is maximized by setting it to a random walk. The paper proves that for any given difference stationary time series, there always exists a decomposition into the sum of a series that is arbitrarily smooth (i.e. "close" to being deterministic) and a stationary residual series. The "long run effect" of a disturbance in the permanent component is shown to be the same regardless of the researcher's assumptions regarding lag lengths and orthogonality between the permanent and transitory components. The theoretical results are applied to examine possible permanent and transitory components in US aggregate output.

1. Introduction.

What is the relative importance of disturbances that are transitory versus disturbances that are permanent in economic fluctuations? What are the dynamic effects of such disturbances? Both empirical and theoretical papers have recently emphasized that these are questions of considerable importance in macroeconomic analysis. A partial list of such papers includes Beveridge and Nelson (1981), Campbell and Mankiw (1987), Clark (1987), Cochrane (1988), Diebold and Rudebusch (1988), Harvey (1985), King, Plosser, Stock and Watson (1987), Nelson and Plosser (1982), Watson (1986) and West (1988).

Suppose that aggregate output does contain a unit root.

Under this maintained hypothesis, one may choose to identify the innovation in aggregate output as the fundamental disturbance to an economy (where innovation is used in the technical time series sense of projection residual on lagged values of the variable itself). If one adopts this view, there is then little left to be said or done: estimating the dynamic response of the economy to this one fundamental shock is simply an exercise in parametrizing the Wold representation.

Alternatively still maintaining the hypothesized unit root property for output, one may conjecture that there is in fact more than one fundamental disturbance that drives aggregate output. It is then an interesting question to disentangle the dynamic effects of the different disturbances. A convenient place to begin is to decompose a unit root time series (aggregate output say) in terms of two components, one permanent and one transitory.

Some well-known such decompositions are the Beveridge-Nelson decomposition (Beveridge and Nelson (1981)), and the unobserved components representation, as in Watson (1986). A key characteristic in these models is that the permanent component is restricted to be a pure random walk, that is, its first difference is serially uncorrelated.

King, Plosser, Stock and Watson (1987), Blanchard and Quah (1988) and Shapiro and Watson (1988) have recently considered models of aggregate output where the serial correlation properties of the permanent component are unrestricted. The motivation for doing this is to study the dynamic effects of those disturbances that will turn out eventually to have permanent impact. Thus this breaks the artificial distinction between short run and long run fluctuations. King, Plosser, Stock and Watson use a common trends rep-

resentation whereas Blanchard and Quah and Shapiro and Watson construct an orthogonal decomposition of output into "interpretable" permanent and transitory components. In the first paper (as in Beveridge and Nelson), the innovations in the common trends are allowed to be correlated with the innovations in the residual terms so that short run and long run fluctuations are not distinct. Thus while each common trend is written as a random walk, its first difference is correlated with the stationary component; strictly speaking therefore the common trend itself is not the entire permanent component. On the other hand, the second and third papers explicitly construct the moving average representation of the permanent component. In both of these cases therefore (i.e. both the orthogonal decompositions in Blanchard and Quah and in Shapiro and Watson, and in the common trends representations in King, Plosser, Stock and Watson) one can meaningfully discuss the dynamic effects of both permanent disturbances and transitory disturbances. By contrast, restricting the permanent component to be a random walk orthogonal to the transitory component simply assumes away any interesting answer to this question.

This paper considers the effects of certain kinds of identifying assumptions for the question of the relative importance of permanent and transitory components. More specifically, this paper uses only orthogonality and lag length restrictions to develop theoretical bounds on the relative magnitudes of permanent versus transitory components. Thus formally the results below relate specifically to the issue of econometric identification.

It is important to carry out the kind of "sensitivity analysis" exercises as in this paper. For example, identifying permanent and transitory components from a bivariate VAR (as in Blanchard and Quah) tempts other researchers to "try a different variable, and see what happens". The results below provide tight (i.e. achievable) bounds on how much these results on the relative importance of permanent and transitory components can change should those other researchers work sufficiently hard. An interesting paper with somewhat similar concerns is Hasbrouck (1988).

The remainder of this paper is organized as follows: Section 2 first provides a general existence proposition for the decomposition of a unit root process into (orthogonal) permanent and transitory components. This section then presents a general approximation result: given an (almost) arbitrary pattern of serial correlation and an (almost) arbitrary observed unit root process, there exists a permanent component for that

process with exactly that pattern of serial correlation. However regardless of the orthogonality assumption between the permanent and transitory components or the hypothesized serial correlation in the permanent component, the long run effect of an innovation in the permanent component is always the same. The results in this section imply that (except in unrealistic degenerate cases) economic examples of difference stationary sequences can be regarded as made up of a stationary part and a "permanent" part that is close to a deterministic time trend. Section 3 specializes the general theory and provides exact calculations for permanent components that are finite ARIMA processes. Let $S(0)$ be the spectral density at frequency zero of the first difference of the observable original process. Then if the first difference of its permanent component is hypothesized to be a q-th order moving average, its variance has greatest lower bound $(q+1)^{-1}S(0)$, while the variance of its innovation has greatest lower bound given by $4^{-q}S(0)$. The greatest lower bound in the autoregressive case is trivial and equal to zero. Section 4 presents the application of the ideas of the preceding sections to US GNP. The paper concludes with a brief Section 5; the Technical Appendix contains proofs of all the results.

2. General Results.

Consider a representative random process Y, assumed to be difference stationary. We wish to view this as being comprised of two kinds of disturbances, one that has permanent effects, and the other having only transitory effects.

Let W be a stochastic sequence and let ΔW denote the first difference sequence: $\Delta W(t) \stackrel{\text{def}}{=} W(t) - W(t-1)$. The elements of a sequence (stochastic or otherwise) will be denoted by integer arguments in parentheses; subscripts will indicate distinct sequences or the elements of a matrix. Thus for example Y_∞ and Y_1 are different stochastic processes, with the t-th element of each written as $Y_\infty(t)$ and $Y_1(t)$. Since there is some arbitrariness in a 2π normalization, we specify explicitly the spectral density matrix to be the fourier transform of the covariogram matrix sequence: for W a covariance stationary vector process, $S_W(\omega) \stackrel{\text{def}}{=} \sum_{j=-\infty}^{\infty} E\left[W(j)W(0)'\right] e^{-i\omega j}$. Unless stated otherwise, all integrals are taken from $-\pi$ to π. All proofs are in the Technical Appendix.

Definition 2.1: *Let Y be a difference stationary sequence. A permanent-transitory (PT) decomposition for Y is a pair of stochastic processes Y_∞, Y_1 such that:*

(i) *Y_∞ is difference stationary, and Y_1 is covariance stationary;*

(ii) $\mathrm{Var}(\Delta Y_\infty), \mathrm{Var}(\Delta Y_1) > 0$;

(iii) $\Delta Y(t) = \Delta Y_\infty(t) + \Delta Y_1(t)$ *in mean square, i.e.,* $E\left[|\Delta Y(t) - \Delta Y_\infty(t) - \Delta Y_1(t)|^2\right] = 0$.

Further:

(iv) If ΔY_∞ is orthogonal to Y_1 at all leads and lags, then the PT decomposition is said to be orthogonal.

Notice that the decomposition of interest is in the sense of mean square (condition *(iii)*): the two stochastic sequences ΔY and $\Delta Y_\infty + \Delta Y_1$ should be indistinguishable in that for each t the difference $\Delta Y(t) - \Delta Y_\infty(t) - \Delta Y_1(t)$ is a random variable with zero mean and variance.

To see why this is important consider the following (incorrect but frequently used) argument: (a.) ΔY_1 is the first difference of a covariance stationary sequence, thus its spectral density vanishes at frequency zero. If further (b.) ΔY_∞ and ΔY_1 are orthogonal at all leads and lags, then the spectral density of their sum is the sum of the individual spectral densities. Then (c.) one can always construct an orthogonal decomposition

of ΔY into ΔY_∞ and ΔY_1: simply let the spectral densities of ΔY and ΔY_∞ be equal at frequency zero. Let the spectral density of ΔY_∞ nowhere exceed that of ΔY, and choose Y_1 so that its first difference has spectral density equal to the difference in spectral densities of ΔY and ΔY_∞. The permanent component ΔY_∞ is of course then arbitrary (so the argument goes) up to satisfying these two conditions, and therefore one can choose ΔY_∞ to have as small a variance as one wishes.

Why is the preceding argument incorrect? The key observation here is that technically all that one has done by the above is simply to write the spectral density of ΔY as the sum of two spectral densities. However there is no sense in which by doing so, a decomposition of Y has actually been constructed. To see this suppose for instance that the variance of the growth rate in US GNP is 1.5 while the variance of the growth rate in consumption of durables is 1.0. Suppose further that the variance of the range in daily temperature as a percentage of the mean temperature on a representative Malaysian beach is 0.5. Is there any sense in which US GNP is the sum of durables consumption and daily temperature? One has not constructed a decomposition of a random process (say of GNP into permanent and transitory components) if one has simply shown an appropriate "adding up" property for second moments.

Notice that this is also why the Wold decomposition theorem is proven by constructing a sequence of projections, and not simply as a result of factoring spectral densities.

This discussion suggests that one should be suspicious of the common practice of simply writing down the decomposition $Y(t) = Y_\infty(t) + Y_1(t)$ without first establishing a general existence result. Beveridge and Nelson have shown that if Y_∞ is a random walk and is perfectly correlated with Y_1, then such a decomposition always exists. It is also well-known however that if Y_∞ is required to be a random walk and orthogonal to Y_1 at all leads and lags, such a decomposition may *not* exist. We therefore provide here a general existence proposition.

To rule out trivial degenerate cases, both permanent and transitory components are required to have strictly positive variances. Notice that the permanent component Y_∞ is not required to be a random walk.

Nelson and Plosser (1982, pp.155-158) briefly treat a first order moving average model for ΔY_∞ in a discussion to bound the relative importance of permanent and transitory components in output. They concluded that for US GNP the standard deviation of innovations in the permanent component relative to

that of the transitory component is in the neighborhood of five or six. The results below may be viewed as generalizing their calculations.

Sometimes it is of interest to impose additional conditions on a PT decomposition. One such set of conditions has been proposed by Blanchard and Quah (1988). Recall that a white noise vector process η is said to be *fundamental* for a covariance stationary vector process W if each entry in η can be recovered as a linear combination of square summable linear combinations of the components of current and lagged values of W. (See for example Rozanov (1965); this is simply the multivariate analogue of Box-Jenkins' invertibility.) For Y a difference stationary sequence, we will identify its innovation with the innovation of its first difference.

Proposition 2.2 (Blanchard-Quah): *Let Y be a difference stationary stochastic process, and let X be such that $(\Delta Y, X)'$ is jointly covariance stationary, and of full rank spectral density. Then there exists an orthogonal PT decomposition (Y_∞, Y_1) for Y such that the vector of innovations in Y_∞ and Y_1 is fundamental for $(\Delta Y, X)$ if and only if ΔY is not Granger causally prior to X. Further, if such an orthogonal PT decomposition exists, then it is unique.*

Thus it turns out that there is an interesting relation between Granger causality and the decomposition proposed by Blanchard and Quah who take Y to be the log of GNP and X to be the measured unemployment rate. In particular by the proof of the Proposition, selecting an X such that it is Granger caused by ΔY, but not vice versa will result in a decomposition that places minimum importance on the transitory component. Blanchard and Quah argue persuasively however that their choice for X is well-justified. (See in particular their discussion of multiple demand disturbances, as well as the results in the Appendix of their paper.) The decomposition asserted in this Proposition will serve as a useful benchmark for the subsequent discussion.

Notice that in the decomposition above, information on the second variable X is crucial for identifying the permanent and transitory components in Y. However there is sometimes interest in isolating the permanent component without using such multivariate information. This may arise out of a suspicion on the part of a researcher that there is actually some well-defined permanent component in Y independent of all other series. The above Proposition indicates that there may be no loss in doing so if and only if Y is Granger causally prior to all other series.

The next proposition provides necessary and sufficient conditions characterizing PT decompositions.

Proposition 2.3: *Suppose that Y is difference stationary, and that Y_∞ and Y_1 are stochastic processes such that $E\left|\Delta Y(t) - \Delta Y_\infty(t) - \Delta Y_1(t)\right|^2 = 0$ for all t. Suppose further that $(\Delta Y_\infty, \Delta Y)'$ is jointly covariance stationary with bounded spectral density matrix $S = \begin{pmatrix} S_{\Delta Y_\infty} & \cdot \\ S_{\Delta Y \Delta Y_\infty} & S_{\Delta Y} \end{pmatrix}$. Then:*

(i) (Y_∞, Y_1) *is a PT decomposition if and only if*

(a) $S_{\Delta Y_\infty}(\omega) = S_{\Delta Y}(\omega) = S_{\Delta Y \Delta Y_\infty}(\omega)$ at $\omega = 0$; and

(b) $\int S_{\Delta Y_\infty}(\omega) d\omega > 0, \quad \int \left[S_{\Delta Y_\infty}(\omega) + S_{\Delta Y}(\omega)\right] d\omega > 2\mathrm{Re} \int S_{\Delta Y \Delta Y_\infty}(\omega) d\omega.$

(ii) (Y_∞, Y_1) *is an orthogonal PT decomposition if and only if*

(a) *as in (i);*

(b) *as in (i);* and

(c) $S_{\Delta Y} \geq S_{\Delta Y_\infty} = S_{\Delta Y \Delta Y_\infty}$ at all ω.

In the following, we will repeatedly use the above alternative representation result to establish that certain proposed candidates are in fact PT decompositions. We also immediately have the following convenient implication:

Corollary 2.4: *If (Y_∞, Y_1) is an orthogonal PT decomposition for a difference stationary sequence Y, then ΔY_∞ is Granger causally prior to ΔY. If further ΔY_∞ and ΔY do not have precisely the same serial correlation pattern, then ΔY is not Granger causally prior to ΔY_∞.*

Thus except in degenerate cases the original series and a permanent component are distinguished by the Granger causality patterns between them.

We now use these characterizations to derive restrictions on the dynamics of possible PT decompositions. The following is the principal result of this paper.

Theorem 2.5: *Let Y be a difference stationary stochastic process, and let (Y_∞, Y_1) be a PT decomposition for Y.*

(i) Suppose $(Y_\infty, Y_1)'$ has a full rank variance covariance matrix, and spectral density matrix strictly positive definite and bounded from above. Let S be a spectral density such that at $\omega = 0$ $S(\omega) = S_{\Delta Y}(\omega)$, and

$\int S(\omega)d\omega > 0$, but is otherwise arbitrary. Then there exists a PT decomposition (X_∞, X_1) for Y such that ΔX_∞ has spectral density equal to S.

(ii) Suppose that (Y_∞, Y_1) is an orthogonal PT decomposition for Y, and let S be a spectral density such that $S_{\Delta Y} - S \geq 0$, with equality at $\omega = 0$, but is otherwise arbitrary. Then there exists an orthogonal PT decomposition (X_∞, X_1) for Y such that ΔX_∞ has spectral density equal to S.

This Theorem is a general possibility result. In words, it says that for any hypothesized serial correlation behavior, there exists a permanent component that has exactly that dynamic pattern of correlations, and such that the deviation from it in the observed Y is covariance stationary.

Notice that the existence claim is proven by explicitly constructing the PT (mean square) decomposition, and therefore circumvents the criticisms above.

The Theorem (together with the preceding Propositions) also makes the following strong identification statement: *regardless of the dynamic structure of the hypothesized permanent component, the long run effect of an innovation in the permanent component is always the same and equal to the square root of the spectral density at frequency zero of the observed data itself*. This extends Watson's (1986) statement for unobserved components models (where the permanent component is restricted to be a random walk and orthogonal to the transitory component) to the case of general permanent-transitory decomposition models, where neither orthogonality nor random walk behavior is assumed.

The results imply that there are many PT decompositions all of whom fit equally well. While they all imply identical long run conclusions (and in fact conclusions that can be adduced without ever estimating any one PT decomposition), each also provides a *different* picture of the short run dynamics implied by a permanent disturbance.

We have the following immediate corollary:

Proposition 2.6: *Let Y be a difference stationary stochastic process, and let (Y_∞, Y_1) be a PT decomposition for Y. Then for any real number $\delta > 0$, there exists a PT decomposition (X_∞, X_1) for Y where* $\mathrm{Var}(\Delta X_\infty) < \delta$.

The implications of this Proposition will be made more concrete in the next section. Notice that this

result applies whether or not one seeks only orthogonal PT decompositions. In words, it says that given any nonstationary process, one can always imagine it to be comprised of a permanent and a transitory component, where the permanent component is arbitrarily smooth (i.e. have the variance of its changes arbitrarily small).

But the limiting case (which is never attained, but can be approached arbitrarily closely) is simply a deterministic trend. Thus this result provides a sense in which those who have argued that difference stationary stochastic processes aren't that different from trend stationary processes are correct. However it differs significantly from other arguments that have been offered in the literature (see for example Clark (1988), Diebold and Rudebusch (1988) or West (1988)) as it imposes *no* requirements on the dynamics of the observed process Y itself, but rather applies regardless of those dynamics.

Cochrane (1987) has also presented an approximation argument that may at first appear identical to the result here. Notice however that (as he correctly emphasizes) his argument is actually one of matching a *finite* number of covariogram terms, and speaks to the problem of econometrically distinguishing difference and trend stationary models when one only has available a finite data segment. By contrast the representation result here is one that applies to the underlying probability model, and is not a problem of statistical inference.

The Proposition on arbitrarily small variance makes no assumptions about the form of the serial correlation permitted. The next section provides exact results when the permanent component is required to be a finite ARIMA process. It turns out this imposes a stricter lower bound on the variance of the first difference than the trivial bound of zero: however the flavor of the current result carries over to that case as well. The results of the next section will also put in perspective the results that have been obtained when Y_∞ is restricted to be a random walk.

3. Finite ARIMA Components.

We first consider finite moving average models for ΔY_∞, and then finite autoregressive models. The results for mixed moving average autoregressive models follow from the finite autoregressive case.

Let $innov(W)$ denote the innovation in the stochastic process W. As above, if W is difference stationary, we identify $innov(W)$ with $innov(\Delta W)$.

Proposition 3.1: *Suppose (Y_∞, Y_1) is a PT decomposition for the difference stationary sequence Y, and ΔY_∞ is a moving average process of order q. Then for $S_{\Delta Y}$ the spectral density of ΔY:*

(i) $\mathrm{Var}(innov(\Delta Y_\infty)) \geq 4^{-q} \cdot S_{\Delta Y}(0)$; and

(ii) $\mathrm{Var}(\Delta Y_\infty) \geq (q+1)^{-1} \cdot S_{\Delta Y}(0)$.

Further, there always exist (different) PT decompositions for Y that have permanent components whose first differences are moving average process of order q, and whose innovation variances and variances are arbitrarily close to the theoretical lower bounds in (i) and (ii).

The lower bounds in Proposition 3.1 are strictly decreasing in the order of the moving average process permitted on ΔY_∞. Thus, we can also immediately conclude that letting ΔY_∞ be a pure random walk maximizes the contribution of the permanent component to Y, in the sense of variance decomposition. The random walk specification sets $q = 1$, and consequently identifies the variance of the change in the permanent component with its innovation variance with the square of the long-run impact of a unit innovation (the sum of the coefficients).

With the result for finite moving average models in hand, the situation for autoregressive models for ΔY_∞ is simple. A first order autoregressive model for ΔY_∞ suffices to obtain an (arbitrarily closely achievable) theoretical lower bound of zero on both its innovation variance and variance. To see this, apply the same arguments as in the proof of the Proposition above to $S_{\Delta Y_\infty}(0) = |1 - C(1)|^{-2} \cdot \mathrm{Var}(innov(\Delta Y_\infty)) = S_{\Delta Y}(0)$, and to $\mathrm{Var}(\Delta Y_\infty) = \left|\frac{1-C(1)}{1+C(1)}\right| \cdot S_{\Delta Y}(0)$, where now $C(1)$ is the projection coefficient in a first order autoregression. Then simply let $C(1) \uparrow 1$. Since a first order autoregressive model already comes arbitrarily close to the trivial lower bound of zero, the same will be true of higher order autoregressive models.

Next, since a purely autoregressive model is simply a restriction of a mixed moving average autoregressive

model, the result for a first order autoregression applies directly to general ARMA models for ΔY_∞.

4. An Empirical Application: GNP.

This section describes the results of applying the ideas of the preceding sections to examining permanent and transitory components in US GNP. First we establish that there exists an orthogonal PT decomposition for aggregate output. From Proposition 2.2, it suffices to find some stationary series X such that the growth rate of aggregate output is not Granger causally prior to X.

Blanchard and Quah (1988) used the measured rate of aggregate unemployment in their study; it is convenient to do so here as well (although any such stationary series will do). Marginal significance levels for testing the coefficients on unemployment to be zero in the projection of output growth on itself and unemployment lagged are 0.86% , 1.71% , and 4.57% for the 4-, 8- and 12-lag bivariate projections. The data are quarterly from 1948:1 to 1987:2 and the marginal significance levels are for the F-statistics reported by the RATS econometric package.

The discussion following Definition 2.1 warns against simply examining the spectral density of output growth for evidence on a permanent-transitory decomposition for output. However the Granger causality results above together with Propositions 2.2 and 2.3 and Theorem 2.5 assure us that in this case, we are not misled by doing so. Figure 4.1 graphs two different estimates of the spectral density function for output growth. One is a smoothed periodogram estimate (using a rectangular two-sided filter of length 17), the other is an autoregressive spectral density estimate. For the second, the eight-lag autoregressive representation is estimated by least squares, then the reciprocal of the square of the fourier transform of the projection representation is graphed here. Under standard regularity conditions, both of these are pointwise consistent for the true spectral density function (see for example Brillinger Theorems 5.6.1 and 5.6.2 and the surrounding discussion [1981, pp.147-9] and Berk Theorem 1 [1974]). Notice that the overall shape of the spectral density estimates in Figure 4.1 are roughly the same, although differing in details.

Our focus here is not the value of the spectral density at any particular fixed frequency, but whether PT decompositions such as those described in Propositions 2.6 and 3.1 can be found for aggregate output. Recall that these decompositions are such that the permanent component is smooth in the sense of having

Figure 4.1

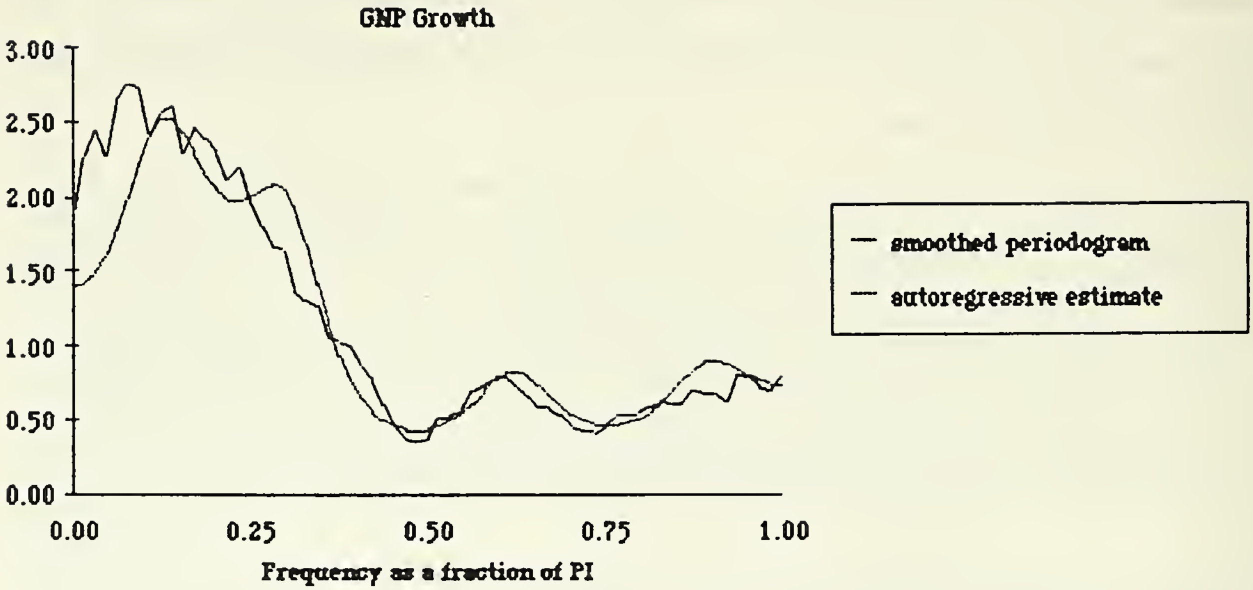

GNP Growth
3.00
2.50
2.00
1.50
1.00
0.50
0.00
0.00
0.25
0.50
0.75
1.00
Frequency as a fraction of PI
smoothed periodogram
autoregressive estimate

Figure 4.2

GNP and Supply Disturbances

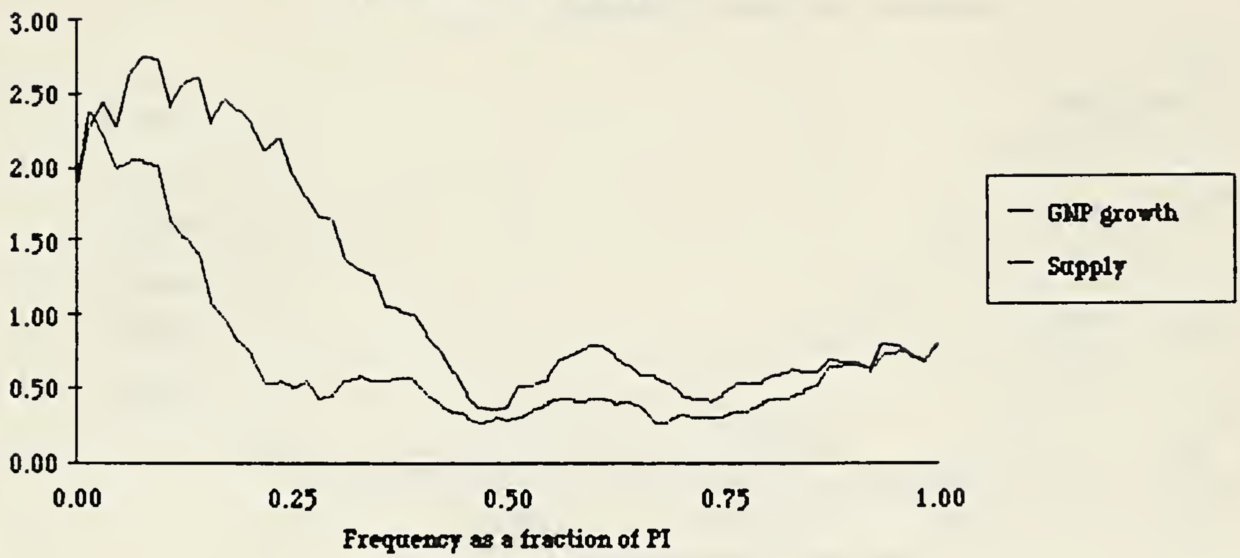


Figure 4.3

GNP Growth, Min Innovation Variance

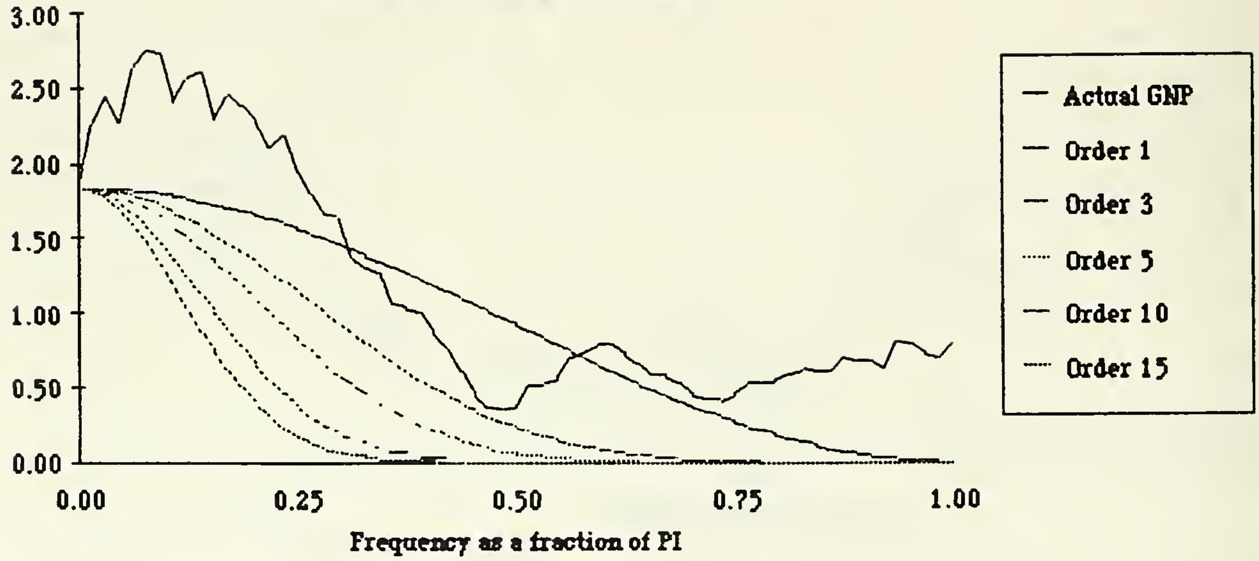
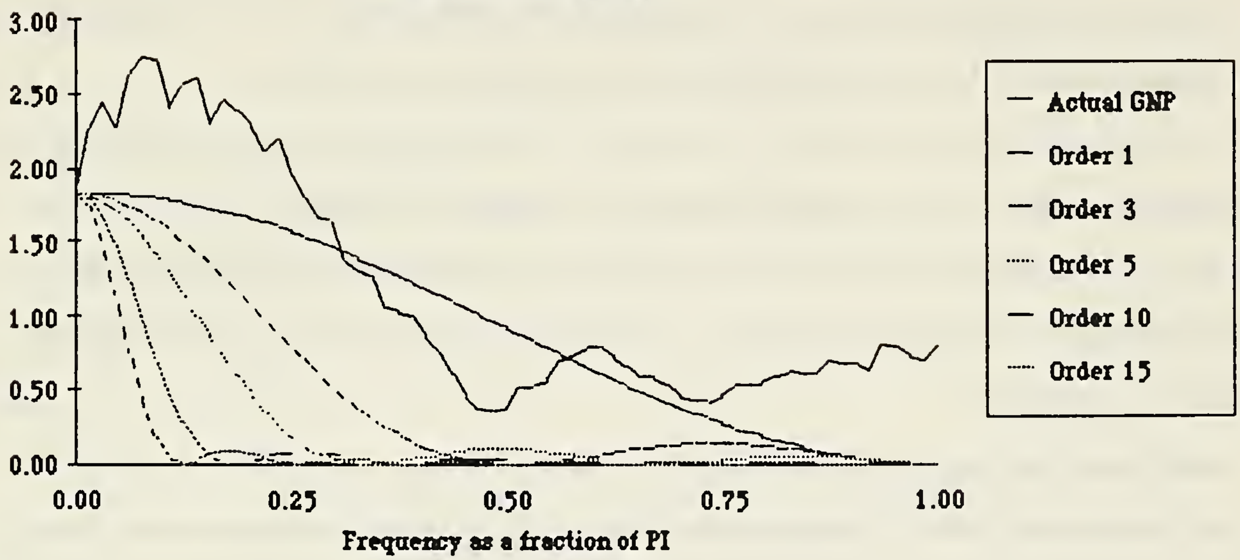


Figure 4.4

GNP Growth, Min Variance



"small" variance.

Figure 4.2 presents the estimated spectral densities for the growth in output and in the supply component of output (in the terminology of Blanchard and Quah). These are smoothed periodogram estimates (unless specified otherwise, all spectral density estimates hereafter are obtained by smoothing the periodogram using a rectangular two-sided window of length 17). The supply component is calculated from "historical realization" by Blanchard and Quah. Due to sampling error, the estimates of the spectral densities at frequency zero of these two stochastic sequences are not exactly equal: they are 1.83 and 1.76 for the original data and for the supply component respectively. For the purposes of graphing the spectral densities in Figure 4.2, that for the supply component is scaled upwards so that the spectral densities coincide at zero. Notice that even in the presence of sampling error and after upwards scaling, except for one or two ordinates the estimate for the supply component never exceeds that for the original data. This is an implication of the orthogonal nature of the supply-demand decomposition in Blanchard and Quah.

The supply component as calculated by those authors is evidently not special in any way, and is certainly not "trivially implied by their assumptions" (I have heard this assertion made a number of times). The results that they actually obtain derive precisely from their use of the unemployment rate as the additional indicator in the system that they estimate. The use of the unemployment rate series is sensible for reasons that are described in their paper.

Figures 4.3 and 4.4 again graph the estimated spectral density function for output growth. Superimposed on this and scaled to coincide at frequency zero are the theoretical spectral densities of moving average permanent components attaining the lower bounds described in Proposition 3.1. Figure 4.3 graphs the spectral densities of minimum innovation variance permanent components (part *(i)* of the Proposition), and Figure 4.4 graphs those of minimum variance permanent components (part *(ii)* of the Proposition).

Without explicitly developing precision properties for these estimates, it is difficult to say if an orthogonal PT decomposition where the permanent component is a pure random walk (say) "fits" aggregate output. The appropriate condition in these graphs would be that the spectral density at zero must also be the minimum value for the spectral density everywhere. However we note that if *some* orthogonal PT decomposition exists, then there necessarily also exists another such decomposition with a permanent component that is

even smoother (in the sense of having a smaller innovation variance or variance): this follows from the way in which richer moving average structures collapse in towards the horizontal axis in Figures 4.3 and 4.4.

5. Conclusion.

This paper has considered the general problem of decomposing a difference stationary process into the sum of a permanent and a transitory component.

It is by now well-known that unit root and trend stationary time series data generate drastically different implications for classical econometric inference. How does this difference carry over onto the observable dynamics of economic variables?

We have shown that without lag length restrictions, the permanent component may be arbitrarily smooth in the sense of having its changes be of arbitrarily small variance. Thus there is a sense in which the observable dynamics in a unit root sequence is close to that in a trend stationary sequence. The precise "long run effect" of a disturbance in the permanent component is always identified and identical, regardless of lag length and orthogonality assumptions.

We have also derived exact lower bounds on the variability in the permanent component when that permanent component is restricted to be a finite ARIMA process. We have shown that the case when the permanent component is a random walk maximizes the importance of that permanent component for explaining the observed data.

In application to US aggregate output, the theoretical results here indicate that GNP can be interpreted as the sum of a stationary component and a permanent component that is arbitrarily smooth. The supply component that is calculated by Blanchard and Quah is seen to be one of many possible permanent components: it is neither the smoothest nor the most volatile.

Technical Appendix.

Proof of Proposition 2.2: *By the Wold Decomposition Theorem,* $\begin{pmatrix} \Delta Y \\ X \end{pmatrix}$ *has a unique moving average representation* $\begin{pmatrix} \Delta Y \\ X \end{pmatrix} = C * \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$, *where* C *is an array of square-summable sequences, zero except on the non-negative integers;* $C(0)$ *is lower triangular;* $*$ *denotes convolution; and* $\epsilon \stackrel{\text{def}}{=} (\epsilon_1, \epsilon_2)'$ *is serially uncorrelated with the identity covariance matrix, and is fundamental for* $(\Delta Y, X)'$. *There exists a unique orthogonal matrix* V *such that* $D \stackrel{\text{def}}{=} CV$ *has its (1,2) entry sum to zero. Writing*

$$\begin{pmatrix} \Delta Y \\ X \end{pmatrix} = D * V'\epsilon = D * \eta, \qquad \text{where } \eta \stackrel{\text{def}}{=} V'\epsilon,$$

we see that η *is fundamental for* $(\Delta Y, X)'$, *and is serially uncorrelated with variance covariance matrix equal to the identity. By the construction, such a* (D, η) *pair is unique, i.e., no other pair admits simultaneously a (1,2) entry in the array of moving average coefficients that sums to zero, and a fundamental disturbance vector that is contemporaneously uncorrelated. Identify* ΔY_∞ *to be* $D_{11} * \eta_1$, ΔY_1 *to be* $D_{12} * \eta_2$. *Since* $\sum_j D_{12}(j) = 0$, ΔY_1 *has spectral density that vanishes at frequency zero, so that then* Y_1 *itself can be chosen to be covariance stationary. Suppose then that* ΔY *is Granger causally prior to* X. *This implies that there exists a moving average representation* $(\Delta Y, X)' = B * \nu$, *where the (1,2) entry in* B *can be taken to be identically zero; the variance covariance matrix of the serially uncorrelated* ν *is arbitrary. It then follows that the (1,2) entry in* C *is identically zero, so that the orthogonal matrix* V *above is simply the identity. But then* ΔY_1 *is identically zero so that* $\mathrm{Var}(Y_1) = 0$. *Thus if* ΔY *is Granger causally prior to* X, *a PT decomposition does not exist. Next suppose the opposite, i.e., consider when* ΔY *is not Granger causally prior to* X. *Then* C_{12} *cannot be identically zero and thus no nontrivial linear combination of* C_{11} *and* C_{12} *is identically zero. It then follows that both* ΔY_∞ *(or equivalently* $D_{11} * \eta_1$*) and* ΔY_1 *(or* $D_{12} * \eta_2$*) have strictly positive variances, and thus are uniquely determined by the above construction.* *Q.E.D.*

Proof of Proposition 2.3: *(i) Suppose* (Y_∞, Y_1) *is a PT decomposition for* Y. *Since* Y_1 *is covariance stationary,* $S_{\Delta Y_1}(\omega) = 0$ *at* $\omega = 0$. *By the inequality* $|S_{\Delta Y_1 \Delta Y_\infty}|^2 \leq S_{\Delta Y_1} \cdot S_{\Delta Y_\infty}$, *this implies that at* $\omega = 0$, $S_{\Delta Y_1 \Delta Y_\infty} = S_{\Delta Y \Delta Y_\infty} - S_{\Delta Y_\infty} = 0$. *Next, recall that* $S_{\Delta Y_1} = S_{\Delta Y_\infty} + S_{\Delta Y} - 2\mathrm{Re} S_{\Delta Y \Delta Y_\infty}$. *At* $\omega = 0$, *this then becomes* $S_{\Delta Y} - S_{\Delta Y_\infty} = 0$. *Thus we have established (a). Further since* $\mathrm{Var}(\Delta Y_1) > 0$,

$$\mathrm{Var}(\Delta Y_1) = \int S_{\Delta Y_1} d\omega = \int [S_{\Delta Y_\infty}(\omega) + S_{\Delta Y}(\omega) - 2\mathrm{Re} S_{\Delta Y \Delta Y_\infty}(\omega)]\, d\omega > 0.$$

Next $\text{Var}(\Delta Y_\infty) = \int S_{\Delta Y_\infty}(\omega) > 0$, *and so we have established (b). To prove the converse, suppose (a). Then at* $\omega = 0$, $S_{\Delta Y_1} = 2S_{\Delta Y_\infty} - 2\text{Re}S_{\Delta Y \Delta Y_\infty} = 2\left[S_{\Delta Y_\infty} - S_{\Delta Y_\infty}\right] = 0$. *Thus* Y_1 *can be chosen to be covariance stationary. By (b), both* ΔY_1 *and* ΔY_∞ *have strictly positive variances. Thus* (Y_∞, Y_1) *is a PT decomposition for* Y. *(ii) Suppose* (Y_∞, Y_1) *is an orthogonal PT decomposition for* Y. *It only remains to establish (c). By orthogonality, for all* ω, $S_{\Delta Y_1 \Delta Y_\infty} = S_{\Delta Y \Delta Y_\infty} - S_{\Delta Y_\infty} = 0$, *which implies that* $S_{\Delta Y \Delta Y_\infty} = S_{\Delta Y_\infty}$. *Further since for all* ω, $S_{\Delta Y_1} = S_{\Delta Y} - S_{\Delta Y_\infty} \geq 0$, *(c.) follows. Conversely, suppose (a) and (b). By (i)* (Y_∞, Y_1) *is a PT decomposition. In addition, if (c) is true, then* ΔY_1 *is orthogonal to* ΔY_∞ *at all leads and lags, so that* (Y_∞, Y_1) *is an orthogonal decomposition.* Q.E.D.

Proof of Corollary 2.4: *By Proposition 2.3,* (Y_∞, Y_1) *being an orthogonal decomposition for* Y *implies that the joint spectral density of* $(Y_\infty, Y_1)'$ *can be written as* $S_{\Delta Y_\infty}\begin{pmatrix} 1 & 1 \\ 1 & 1+\psi \end{pmatrix}$ *for some real symmetric positive definite function* ψ, *zero at frequency zero,* $S_{\Delta Y} = (1+\psi)S_{\Delta Y_\infty} \geq S_{\Delta Y_\infty}$. *But then the projection of* ΔY *on lead, current and lag values of* ΔY_∞ *has coefficients whose fourier transform is* $S_{\Delta Y_\infty}/S_{\Delta Y_\infty} = 1$. *This is simply the identity however and therefore places zero weight on lead values. Thus* ΔY_∞ *is Granger causally prior to* ΔY. *Next the projection of* ΔY_∞ *on lead, current and lag values of* ΔY *has coefficients with fourier transform* $(1+\psi)^{-1}$. *When* ΔY_∞ *and* ΔY *do not have exactly the same serial correlation patterns,* ψ *varies over* $(-\pi, +\pi]$. *But then the coefficient sequence has a fourier transform that is nontrivial and real, and thus the coefficient sequence itself is two-sided and symmetric. Consequently* ΔY *is not Granger-causally prior to* ΔY_∞. Q.E.D.

Proof of Theorem 2.5: *(i) Choose* b *to be the (one-sided) sequence of Wold moving average coefficients for the spectral density* $S/S_{\Delta Y_\infty}$, *i.e.,* $|\tilde{b}|^2 = S_{\Delta Y_\infty}^{-1} S$. *Notice that since* $S(\omega) = S_{\Delta Y}(\omega) = S_{\Delta Y_\infty}(\omega)$ *at* $\omega = 0$, *we have* $\sum_j b(j) = 1$. *Set* $X_\infty = b * Y_\infty$, *and* $X_1 = Y - X_\infty$. *The spectral density matrix of the jointly covariance stationary vector sequence* $(\Delta X_\infty, \Delta Y)'$ *is:*

$$\begin{aligned}\begin{pmatrix} S_{\Delta X_\infty} & \cdot \\ S_{\Delta Y \Delta X_\infty} & S_{\Delta Y} \end{pmatrix} &= \begin{pmatrix} \tilde{b} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} S_{\Delta Y_\infty} & \cdot \\ S_{\Delta Y \Delta Y_\infty} & S_{\Delta Y} \end{pmatrix} \begin{pmatrix} \tilde{b}^* & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} |\tilde{b}|^2 S_{\Delta Y_\infty} & \cdot \\ \tilde{b} S_{\Delta Y \Delta Y_\infty} & S_{\Delta Y_\infty} \end{pmatrix} = \begin{pmatrix} S & \cdot \\ \tilde{b} S_{\Delta Y \Delta Y_\infty} & S_{\Delta Y_\infty} \end{pmatrix}.\end{aligned}$$

Thus ΔX_∞ *has the required dynamics in* S. *Further since* $\tilde{b} = 1$ *at* $\omega = 0$, *and* (Y_∞, Y_1) *is a PT decomposition for* Y, *this matrix is seen to have all its elements equal at* $\omega = 0$. *Finally, the determinant of this spectral*

density matrix is

$$S \cdot S_{\Delta Y} - |\tilde{b}|^2 \, |S_{\Delta Y \Delta Y_\infty}|^2 = |\tilde{b}|^2 \cdot \det \begin{pmatrix} S_{\Delta Y_\infty} & \cdot \\ S_{\Delta Y \Delta Y_\infty} & S_{\Delta Y} \end{pmatrix}$$

which implies:

$$\int \det \begin{pmatrix} S_{\Delta X_\infty} & \cdot \\ S_{\Delta Y \Delta X_\infty} & S_{\Delta Y} \end{pmatrix} (\omega)\, d\omega > 0.$$

Because in addition $\int S(\omega)\, d\omega > 0$*, we then have that the variance covariance matrix of* $(\Delta X_\infty, \Delta Y)'$ *is full rank. By Proposition 2.3,* (X_∞, X_1) *is a PT decomposition for* Y*. (ii) Since* (Y_∞, Y_1) *is an orthogonal PT decomposition for* Y*, Proposition 2.3 implies that the spectral density matrix of* $(\Delta Y_\infty, \Delta Y)'$ *can be written as* $S_{\Delta Y_\infty} \begin{pmatrix} 1 & 1 \\ 1 & 1+\psi \end{pmatrix}$*, for* ψ *some positive definite real symmetric function with* $\psi(\omega) = 0$ *at* $\omega = 0$*. For* b, c *square summable sequences, define* $\Delta X_\infty(b,c)$ *to be the process* $b * \Delta Y_\infty + c * \Delta Y$*. By construction,* $(\Delta X_\infty, \Delta Y)'$ *is jointly covariance stationary, and has spectral density matrix given by:*

$$\begin{pmatrix} \tilde{b} & \tilde{c} \\ 0 & 1 \end{pmatrix} S_{\Delta Y_\infty} \begin{pmatrix} 1 & 1 \\ 1 & 1+\psi \end{pmatrix} \begin{pmatrix} \tilde{b}^* & 0 \\ \tilde{c}^* & 1 \end{pmatrix}.$$

Up to a stochastic process whose first difference vanishes in mean square, $X_\infty(b,c)$ *is uniquely defined by the requirement that its first difference be* $b * \Delta Y_\infty + c * \Delta Y$*. Let* $X_1(b,c) \stackrel{\text{def}}{=} Y - X_\infty(b,c)$*. By Proposition 2.3, for* $(X_\infty(b,c), X_1(b,c))$ *to be an orthogonal PT decomposition, it is necessary and sufficient that there exist symmetric positive definite functions* S_X, ψ_X*, with* $\psi_X(\omega) = 0$ *at* $\omega = 0$*, such that the spectral density matrix of* $(\Delta X_\infty, \Delta Y)$ *above can be represented as* $S_X \begin{pmatrix} 1 & 1 \\ 1 & 1+\psi_X \end{pmatrix}$*. Set* S_X *to* S *in the statement of the Theorem, and define* $\psi_X \stackrel{\text{def}}{=} S_X^{-1} S_{\Delta Y_\infty}(1+\psi) - 1 = S^{-1}(S_{\Delta Y} - S)$*, which is therefore guaranteed to be a real symmetric positive definite function, vanishing at* $\omega = 0$*. Select square summable sequences* b, c *such that their fourier transforms* $\tilde{b}, \tilde{c}$ *satisfy:*

$$\left|\tilde{b}\right|^2 = (S/S_{\Delta Y_\infty})^2 \cdot \psi_X/\psi \qquad \text{for } \omega \neq 0, \text{ and } 0 \text{ at } \omega = 0;$$

$$\tilde{c} = (1+\psi)^{-1} \left[(S/S_{\Delta Y_\infty}) - \tilde{b} \right].$$

Notice that $\tilde{b}$ *is restricted only to the extent that its modulus on each frequency satisfies the above equality. Since the right hand side is real symmetric and positive definite, it is a spectral density. Thus* b *can be chosen to be the (one-sided) Wold moving average coefficients corresponding to that spectral density. It is straightforward to verify that these sequences* b, c *imply that:*

$$\begin{pmatrix} \tilde{b} & \tilde{c} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1+\psi \end{pmatrix} \begin{pmatrix} \tilde{b}^* & 0 \\ \tilde{c}^* & 1 \end{pmatrix} = \left(\frac{S_X}{S_{\Delta Y_\infty}} \right) \begin{pmatrix} 1 & 1 \\ 1 & 1+\psi_X \end{pmatrix}$$

Thus $(X_\infty(b,c), X_1(b,c))$ is an orthogonal PT decomposition. Q.E.D.

Proof of Proposition 2.6: *It is always possible to choose S in the Theorem to be such that $\frac{1}{2\pi}\int S(\omega)d\omega$ is no greater than δ.* Q.E.D.

Proof of Proposition 3.1: *Since ΔY_∞ is a moving average process of order q,*

$$\Delta Y_\infty(t) = \sum_{j=0}^{q} C(j)\, innov(\Delta Y_\infty)(t-j), \qquad \text{for all } t,$$

with $C(0)=1, \sum_{j=0}^{q} C(j)z^j \neq 0$ for $|z|<1$. Since Y_∞ is a permanent component for Y,

$$\left|\sum_{j=0}^{q} C(j)\right|^2 \cdot \mathrm{Var}(innov(\Delta Y_\infty)) = S_{\Delta Y}(0).$$

Thus the lower bound on $\mathrm{Var}(innov(\Delta Y_\infty))$ is obtained by solving:

$$\sup_C \left|\sum_{j=0}^{q} C(j)\right|^2 = \left|\left(\sum_{j=0}^{q} C(j)z^j\right)|_{z=1}\right|^2$$

$$\text{subject to } C(0)=1, \qquad \sum_{j=0}^{q} C(j)z^j \neq 0 \text{ for } |z|<1.$$

Recall that any such polynomial $\sum_{j=0}^{q} C(j)z^j$ above may be written as the product of q monomials: $\sum_{j=0}^{q} C(j)z^j = \prod_{j=1}^{q}(1+D(j)z)$, with $|D(j)| \leq 1, j=1,2,\ldots,q$ appearing in complex conjugate pairs if not real. Since $|\sum C(j)z^j|^2 = |\prod_{j=1}^{q}(1+D(j)z)|^2 = \prod_{j=1}^{q}|1+D(j)z|^2$, its maximization at $z=1$ is equivalent to the maximization of $|1+D(j)|^2$, for each $j=1,2,\ldots,q$. This occurs at $D(j)=1$ for each j. Therefore the solution to the optimization problem attains the value 4^q. The lower bound on the innovation variance is then $4^{-q}\cdot S_{\Delta Y}(0)$, and results when the moving average representation for ΔY_∞ is $(1+L)^q innov(\Delta Y_\infty)$, where L is the lag operator. Next, the lower bound on $\mathrm{Var}(\Delta Y_\infty)$ is obtained by solving:

$$\inf_{(C,\sigma^2)} \sigma^2 \sum_{j=0}^{q} C(j)^2$$

$$\text{subject to } C(0)=1, \sum_{j=0}^{q} C(j)z^j \neq 0 \text{ for } |z|<1, \text{ and } |\sum_{j=0}^{q} C(j)|^2\sigma^2 = S_{\Delta Y}(0).$$

Substituting out for σ^2*, we need to minimize* $\left(\sum_{j=0}^{q} C(j)^2\right) / \left|\sum_{j=0}^{q} C(j)\right|^2$ *subject to the boundary conditions above. First notice that* $\left|\sum_{j=0}^{q} C(j)\right|^2 \leq \left(\sum_{j=0}^{q} |C(j)|\right)^2$*. Next apply the Cauchy-Schwarz inequality:*

$$\left|\sum_{j=0}^{q} C(j)\right|^2 \leq \left(\sum_{j=0}^{q} |C(j) \cdot 1|\right)^2 \leq \left(\sum_{j=0}^{q} |C(j)|^2\right) \cdot \left(\sum_{j=0}^{q} 1^2\right) = \left(\sum_{j=0}^{q} C(j)^2\right) \cdot (q+1).$$

Therefore we have:

$$\left(\sum_{j=0}^{q} C(j)^2\right) / \left|\sum_{j=0}^{q} C(j)\right|^2 \geq (q+1)^{-1}.$$

Notice that $C(j) = 1$ *for* $j = 0, 1, \ldots, q$*, achieves this lower bound, and because* $\sum_{j=0}^{q} z^j = \lim_{\lambda \uparrow 1} \frac{1-\lambda^{q+1} z^{q+1}}{1-\lambda z}$*, this satisfies the boundary conditions as well. Thus* $\mathrm{Var}(\Delta Y_\infty) \geq (q+1)^{-1} S_{\Delta Y}(0)$*, and this theoretical lower bound is evidently approached arbitrarily closely by finite moving average processes of the form* $\sum_{j=0}^{q} \lambda^j \mathrm{innov}(\Delta Y_\infty)(t-j)$*, where* $\lambda \leq 1$*.* *Q.E.D.*

References

Berk, K.N. (1974): "Consistent Autoregressive Spectral Estimates," *Annals of Statistics*, 2 no.3, 489-502.

Beveridge, S. and C.R. Nelson (1981): "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics*, 7, 151-174.

Blanchard, O.J. and D. Quah (1988): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," mimeo, February, MIT.

Brillinger, D. (1981): *Time Series, Data Analysis and Theory*. San Francisco: Holden-Day.

Campbell, J.Y. and N.G. Mankiw (1987): "Are Output Fluctuations Transitory?" *Quarterly Journal of Economics*, November, 857-880.

Clark, P.K. (1987): "The Cyclical Component in US Economic Activity," *Quarterly Journal of Economics*, November, 797-814.

Clark, P.K. (1988): "Nearly Redundant Parameters and Measure of Persistence in Economic Time Series," Graduate School of Business, Stanford University, January.

Cochrane, J. (1987): "A Critique of the Application of Unit Roots Tests," University of Chicago Working Paper, December.

Cochrane, J. (1988): "How Big is the Random Walk Component in GNP?" forthcoming, *Journal of Political Economy*.

Diebold, F.X. and G.D. Rudebusch (1988): "Long Memory and Persistence in Aggregate Output," Finance and Economics Discussion Series, Division of Research and Statistics, Federal Reserve Board, January.

Harvey, A.C. (1985): "Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics*, 3 no. 3, 216-227.

Hasbrouck, J. (1988): "Bounding the Variance of the Stationary Component of a Detrended Univariate Time Series," NYU Department of Finance Working Paper, May.

King, R., C. Plosser, J. Stock and M. Watson (1987): "Stochastic Trends and Economic Fluctuations," NBER Working Paper No. 2229, Cambridge, April.

Nelson, C. (1987): "A Reappraisal of Recent Tests of the Permanent Income Hypothesis," *Journal of Political Economy*, 95 No. 3, 641-646.

Nelson, C. and C. Plosser (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139-162.

Rozanov, Y.A. (1965): *Stationary Random Processes*, San Francisco: Holden-Day.

Shapiro, M. and M. Watson (1988): "Sources of Business Cycle Fluctuations," in Fischer, S. (ed.): *NBER Macroeconomics Annual 1988*, Cambridge: MIT Press.

Watson, M.W. (1986): "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics*, July, 18, 1-27.

West, K.D. (1988): "On the Interpretation of Near Random Walk Behavior in GNP," *American Economic Review*, March, 78 no. 1, 202-209.

3-22-5